

AI consciousness and neuroscientifically plausible “seamless” mind-uploading

openaccessgovernment.org/article/ai-consciousness-and-neuroscientifically-plausible-seamless-mind-uploading/168802

24 October 2023

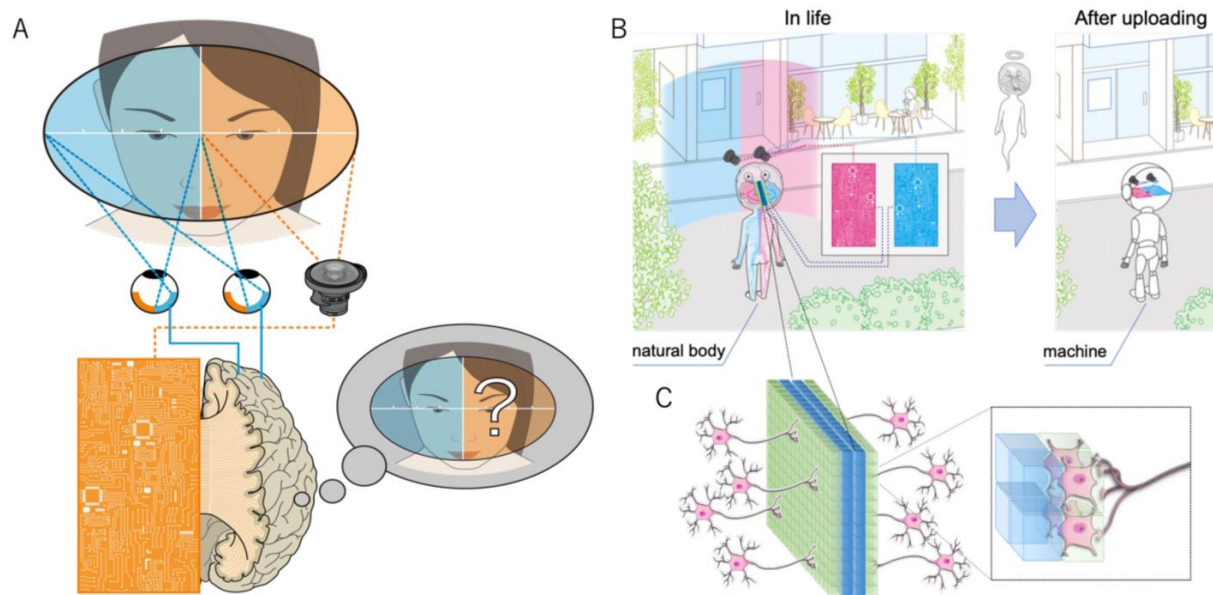


Figure 1. A) A test for AI consciousness B) Integrating AI and biological hemispheres with C) a radically new brain-machine-interface that reads and writes from the surface of dissected axonal bundle (e.g., corpus callosum)

Masataka Watanabe, Associate Professor at the University of Tokyo’s School of Engineering, examines a test for AI consciousness. He proposes it as part of a scientific approach to deciphering consciousness that leads to “seamless” mind uploading

Today, there is much debate on whether artificial intelligence (AI) has achieved consciousness. ^(1, 2) My short answer is probably not, but with dedicated substantial resources, AI will achieve consciousness at some point.

So, what is consciousness? Could we ever tell whether it resides in AI? How can we build a conscious AI, and what would be the purpose of doing so?

Defining consciousness

According to the philosopher Thomas Nagel, “an organism has conscious mental states if and only if there is something that it is like to be that organism – something it is like for the organism.”⁽³⁾ We can be sure that there is an element of what it is like to be a human. We experience vision, audition, and various other sensations through our brain.

What about smartphones equipped with modern AI? When we record videos on our phones, it not only records visual information but also detects faces, etc., on the fly and adjusts its lens for focus. But when all that is happening, is there something to being a “smartphone”? Are they seeing the world as we are? Most probably not.

Difficulty in testing AI consciousness

The difficulty in testing AI consciousness is that we must account for philosophical zombies, a concept proposed by the philosopher David Chalmers. A philosophical zombie is indistinguishable from a normal human being in terms of its appearance or behaviour; the only difference is that a philosophical zombie lacks consciousness.

Philosophical zombies are an imaginary construct, but a necessary one for working out the constraints of testing AI consciousness; it hinders testing with objective measures such as externally observing stimulus responses.

Together, according to the Mill Argument by the philosopher Gottfried Wilhelm Leibniz, it is also impossible to judge by deciphering its inner workings.

Need for a subjective test

If it is impossible to objectively test for AI consciousness, only one method remains: making use of subjectivity. We need to connect our own brains to the machine and “see” for ourselves whether consciousness resides within.

One critical issue is how we connect the machine to our own brains. We need a way to connect the device so that we obtain a subjective sensory experience, and only when consciousness resides in it.

My proposed test for machine consciousness takes advantage of the primary–primary constraint of our visual system (Fig. 1A).^(4,5) As shown by Sperry’s Nobel prize-winning study, with regard to visual consciousness, the left and right hemispheres are equal. Namely, there is no asymmetric relationship in which one hemisphere generates consciousness while the other just provides it with visual information.

The test runs as follows: replace one of our hemispheres with an AI hemisphere to see whether we subjectively experience integrated visual fields, including the side that the mechanical hemisphere processes. If we do, from the primary–primary constraint, we must

conclude that a stream of visual consciousness resides in the AI hemisphere and that it is linked to our own biological stream of consciousness.

Building conscious AI: Mind uploading

Many scientists believe that, in principle, conscious AI is possible; if we manage to sufficiently replicate the neural circuitry and dynamics of a human brain in artificial neural networks, it would be rather mysterious if consciousness does not reside.

But then, what would be the very purpose of developing AI consciousness? If our only goal is to develop an AI that acts like it were conscious, the authenticity of that consciousness is unimportant.

As long as AI behaves as if it is conscious, we don't need to determine whether it genuinely is. However, if our goal is to upload our own mind, things are very different; we certainly wouldn't want to find ourselves transformed into a philosophical zombie after the process.

AI consciousness: A scientifically feasible path to seamless mind uploading

I put forward a three-step procedure to realise AI consciousness and, subsequently, "seamless" mind uploading.

In the first step, a neutrally conscious device is constructed. The idea is to prepare a spiking neural network (SNN) that replicates the full connectivity of the human brain based on data obtained from future invasive connectome projects. From there, to determine the fine quantitative values of neuronal connectivity and develop the device into a visual system, for instance, we can show it a life's worth of video material.

By using advanced methods for training SNNs, the final result should mimic a human brain, and most likely, neutral consciousness, or a 'one-size-fits-all' type of consciousness would likely kick in. The beauty lies in the fact that we may test whether consciousness has actually emerged using the above-proposed test.

Once we know for certain that the AI hemisphere is conscious, we are only one step away from mind uploading. Since the two streams of consciousness – one in AI and the other in the biological hemisphere – would be readily integrated, all we need to do is to transfer memory from the biological to the AI side.

This step includes natural and forced memory retrieval in the biological hemisphere, leading to synchronised retrieval in the artificial hemisphere due to integrated consciousness between the two. Finally, memory consolidation is achieved in the device side with brain-like mechanisms.

Once consciousness is fully integrated and sufficient memories have been transferred, we will be ready to face the inevitable closure of the two biological hemispheres. This closure would be like suffering a massive stroke in one of our hemispheres where our consciousness seamlessly continues in the other hemisphere, except that our consciousness will continue in the two AI hemispheres, to be integrated later (Fig. 1B, C).

References

1. Elizabeth Finkel, "If AI becomes conscious, how will we know?", *Science*, (2023) doi: <https://doi.org/10.1126/science.adk4451>
2. Mariana Lenharo, "If AI becomes conscious: here's how researchers will know", *Nature*, (2023) doi: <https://doi.org/10.1038/d41586-023-02684-5>
3. https://www.liquisearch.com/who_is_thomas_nagel
4. Masataka Watanabe, "From Biological to Artificial Consciousness", Springer, (2022) doi: <https://doi.org/10.1007/978-3-030-91138-6>
5. Masataka Watanabe, Kang Cheng, Yusuke Murayama, Kenichi Ueno, Takeshi Asamizuya, Keiji Tanaka, Nikos Logothetis. Attention But Not Awareness Modulates the BOLD Signal in the Human V1 During Binocular Suppression. *SCIENCE*. 2011. 334. 6057. 829-831 <https://doi.org/10.1126/science.1203161>

Please Note: This is a Commercial Profile



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).